Back to the Future of Green Powered Economies

Juan Moreno Cruz       M. Scott Taylor*

July 8, 2013

Preliminary — Please do not cite or distribute without permission.

Abstract

The purpose of this paper is to examine how the location and productivity of available energy resources affects the spatial distribution of economic activity. More productive (power dense) resources generate far greater energy supply, stronger incentives for infrastructure investments, higher consumption per capita, and more population dense agglomerations. Using county level population data from 1086 to 1841 we investigate how England’s move to the very power dense energy source represented by coal altered its economic geography. Local access to coal is responsible for over 25% of growth in all but one coal county; and more than 50% in others. The world’s first energy transition created a dramatic reshuffling of the economic landscape.

*Moreno-Cruz, Assistant Professor, School of Economics, Georgia Institute of Technology, Atlanta, Georgia; Taylor, Professor of Economics and Canada Research Chair, Department of Economics, University of Calgary, Canada and Research Associate at the NBER. We are very grateful to seminar audiences at several universities, and would like to thank, without implicating, Bob Allen, Herb Emery, Oded Galor, Gene Grossman, Arvind Magesan and Jim Wilen for comments. The usual disclaimer applies. Excellent research assistance was provided by Jevan Cherniwchan, Nolan Derby and Rui Wan.
1 Introduction

In the not so distant past, solar power captured by wind flows and biomass production were the sole energy sources fueling human existence. Students of this era, particularly economic historians, have studied the role these nature given limits played in determining the economic limits to land use, and debate to this day the impact fossil fuels have had on economic activity. Students of our current era, particularly energy economists, look to a not so distant future where post-industrial green powered economies may again survive on the direct and indirect fruits of solar power. And similarly, energy economists today debate the economic effects of moving back to renewables away from fossil fuels. These two disciplines, separated by several hundred years in their respective periods of study, share a common interest in understanding how the introduction of a new energy source affects economic activity.

The purpose of this paper is to take a step back into the past to consider what our future may look like by examining how the introduction of a new energy source affects the size and location of economic activity. To do so we build a new spatial model where energy resources differ in their productivity and location, and then use this model to develop a novel empirical design to evaluate its implications. Our main theoretical result is the statement of the density-creates-density hypothesis which links the power or energy density of neighboring resources with the resulting geographic distribution of population and economic activity. Our main empirical result is the finding that the geographic location of coal deposits in pre-fossil fuel English counties, caused a major redistribution of population and economic activity across these counties once the economy transitioned to coal. All coal counties grew dramatically because of coal; some owe more than 50% of their growth to coal; and access to coal reshaped the entire distribution of people and production. Since our sample period runs from 1086 to 1841 many other forces are also at play in the English countryside over this tumultuous 800 years. Coal is not the driver of all changes, nor perhaps even the majority of changes, but it is a very important driver nonetheless.

To understand how access to energy resources affects the location of production we need a model of economic geography, and any model of economic geography must have three components: a reason why agglomerations exist; a reason why they are bounded in size; and a mechanism for sorting people and activity across locations. Our work is no different. The forces for agglomeration in our model arise for very standard reasons: increasing returns from specialization in production are facilitated by large markets, although the specifics are tailored to our energy application. Agglomerations are limited in size by an explicit spatial structure which ensures a rising cost of bringing marginal energy resources to the core. To
sort people across locations, we adopt a mechanism that fits both the time and place of 
our application - a simple Malthusian mechanism where individuals sort across competing 
locations on the basis of death rates and consumption per capita.

A novel element of our theory is that we differentiate energy resources accordingly to 
their "spatial productivity" and measure this productivity using the concept power density. 
The concept power density is unfamiliar to most economists but it simply measures the flow 
of energy a source can provide per unit area needed for its exploitation. Power is the flow 
of energy per unit time, and this together with the flow of labor and capital services is what 
drives economic activity. While our explicit measure of spatial productivity is rarely used in 
economic analyses, the productivity of soil, the yield of various crops, the regeneration power 
of forests, the energy content of fuels, and the fecundity of animal species are all reflective 
of this more general measure.\(^1\) Seen in this light, measures like power density have received 
considerable attention in the analyses of economic historians where the interaction of an 
environment’s spatial productivity and transport costs are held responsible for the small 
city sizes in pre-industrial times.\(^2\) These arguments are however exclusively verbal, and to 
examine how variation in the productivity of energy resources across space affects locational 
choices we develop an explicit model of energy exploitation, transportation and use.

Our spatial setting allows us to model the costly exploitation and transport of energy 
explicitly. We assume transport is costly because it requires energy, and calculate the work 
needed to move real resources across space. Transport requires both energy and other 
inputs and we assume both inputs are essential. Since energy has to be collected, refined, 
transported, etc. it is difficult to imagine a world where energy could be delivered at zero 
energy cost to consumers; it is also impossible to think of energy being collected, moved 
or distributed without complementary inputs such as wheel barrows, wagons, pipelines or 
transmission lines that comprise the energy system.

To model an energy system we follow both energy economists and economic historians 
alike who agree that every fuel source has unique attributes that often requires - and always 
leads to - the introduction of a set of facilitating goods that aid in its collection, transport, 
and conversion. For example, scythes, carts, and horse collars were all important facilitating 
goods in pre-industrial economies just as today solar collectors, wind turbines, and DC 
converters are important for today’s renewables. Since these goods are “intermediate” in 
the supply chain somewhere between the exploitation and collection, and the final use of

\(^1\)In Moreno-Cruz and Taylor (2013) we link the power density of both renewables and non-renewables to 
these more primitive determinants.

\(^2\)The link between city size and power density is very clearly made in many contributions of Vaclav Smil 
(see for example Smil (2006, 2008)); it is also a recurrent theme in the work of Wrigely (see Wrigley (2010) 
for one example) although Wrigley does not use the term power density. See also Nunn and Qian (2011).
energy, we will refer to them as intermediate goods. The role of these intermediate goods is to facilitate the conversion of raw energy inputs into final energy services at a lower cost than otherwise. The combination of an energy source, its set of facilitating intermediate goods, and the pattern of production and consumption defines, what we will call, an “energy system.” While referring to a set of interrelated demands and supplies as a system is standard fare in economics, energy system experts and energy historians have a more inclusive definition in mind. Their energy system is one where complementarities between and across system components link consumption, transport and exploitation in a mutually reinforcing way creating elements of increasing returns and path dependence. To capture and hopefully understand the potential role played by these system wide complementarities we adopt a framework where increased specialization in the set of intermediate goods increases the efficacy of the energy system.

Finally we must allow for the possibility that cities compete for population. To do so, we couple our market model with a simple model of Malthusian population dynamics where consumption per capita and population density jointly determine birth and death rates. The Malthusian application gives us a simple mechanism to sort population across space and investigate the density-creates-density hypothesis. We then evaluate its predictions by reviewing the population history of England from 1086 to 1841 through the lens of our theory.

Our theory provides four key results. First, even small differences in the power density of resources can create large differences in equilibrium outcomes. Specifically, we show how energy supply is a cubic function of power density. This reflects a scaling law coming from our spatial context. Resources with twice the power density, deliver eight times the energy supply. Since straw and wood are not very power dense resources, this finding confirms the key role economic historians have given to energy constraints in pre-industrial economies. Second, there is a strong complementarity between the power density of energy resources and the incentives we have to provide improved transportation. Resources with low power density - dung, straw, wood - provide meagre incentives for transportation improvement; resources with high power densities - coal and oil - provide much larger incentives. This complementarity magnifies the already large difference in the supplies of more dense and less dense fuels. Third, the interaction of rising marginal costs of energy exploitation and increased efficiencies in the energy delivery system, produce a first rising and then falling schedule of per capita consumption as the size of markets grow. Large markets require marginal and costly resources to fuel them; but they also support the introduction of many complementary inputs which makes the consumption and delivery of energy services less costly. This interplay produces a peak in the consumption per capita profile, and this
peak is higher and at larger market sizes when the resources in question are more power
dense. Finally, if agents sort across locations trading off consumption per person and
expected lifetime then locations with very productive neighboring resources create very dense
agglomerations - this is the density-creates-density hypothesis.

Our model is highly abstract and of very small dimension. It fits no real world economy
past nor present, but does carry with it several potentially testable hypotheses. To move
towards an empirical evaluation of the density-creates-density hypothesis we first revisit the
economic and population history of England from the Norman conquest to the start of the
Victorian era. Using our theory as a lens we review the first order developments of the
day and link them to implications for data. The Norman invasion, the Black Death, the
end of Serfdom, the rise of London and the movement to coal all feature prominently in
our narrative; but most importantly, our discussion suggests a natural empirical strategy for
making a causal argument linking access to coal and changes in the geographic distribution of
population. Our empirical design boils down to a relatively simple difference-in-difference
estimation where over time and cross sectional variation in - the ability to use, and the
fortune to have - coal deposits determines treatment.

Our empirical results suggest access to coal deposits had large effects on the growth of coal
counties. Since the overland transport of coal was difficult and expensive, local abundance
of coal was critical to its supply and price. With the exception of Newcastle coal delivered at
low cost to London, most areas relied on local deposits and faced very local energy markets.
In this very segmented world, we find coal counties often owed 30% or more of their growth
to coal. Of the lucky 13 coal counties, only one had less than 20% of its growth over the
treatment period attributed to coal; several owe coal more than 40%, and one county more
than 50%. While these impacts are large, other forces at work in the English economy
were equally or more powerful over this period. Textile production, booming international
trade, major political developments and the benefits of empire all played a role in reshaping
England over this period. To assess the importance of coal relative to these other forces,
we construct a counterfactual population distribution for English counties in a world where
coal counties do not benefit from their preferential access. This construction shows a more
nuanced interpretation is in order. While having local deposits helped all of the lucky 13
cal counties, five of the coal counties would have moved up the economic hierarchy without
it; three of them were destined to remain laggards in any case; and only five of the coal
counties were truly transformed by their access. Nevertheless, our results strongly support
the density-creates-density hypothesis.

Our work is related to contributions coming from three large and largely disjoint litera-
tures: economic history; energy economics; and economic geography. In Moreno-Cruz and
Taylor (2013) we developed a very simple spatial model to show how to incorporate energy
density concepts into standard economic analysis. We provided methods to measure power
density for both renewables and non-renewables, and examined the role it may play in a very
simple economic system. Since energy was the only input and only output of the production
process, we referred to our simple spatial model as the Only Energy Model. Here we have
effectively embedded our Only Energy Model in a market economy setting. Like Henderson
(1974, 1980) the interplay of increasing returns and transport costs generates an optimal
city size. We differ however by linking the density of economic agglomeration to the density
of the energy sources supplying the city. And similar to Fujita et al (2000) we generate
increasing returns at the economy level from increased specialization in economic activity.

Work in economic history is also related. Perhaps most importantly, Wrigley (2010)
argues the low density of available energy sources in the U.K made urbanization and further
progress impossible in what he refers to as the Organic Economy period. He argues for a
view of the Industrial Revolution where positive feedbacks between the density of fossil fuels,
the resulting urbanization, and eventual technological progress drive the transformation of
the UK economy as it enters the fossil fuel driven Mineral economy period. Similarly, Smil
(2008) argues that the low density of biomass based fuels kept villages small and the market
size for any innovation too low to foster growth. Recent work by Nunn and Qian (2011) is
also directly on point as they link the introduction of a new high density energy source (the
lowly potato) to population growth and urbanization in the Old World. Important work
specific to the period we study is Wrigley (1969, 1998), Allen (2009), and Clark (2009). Our
discussion of English history and the role of energy owes much to Wrigley (2010) and Allen
(2009); our empirical design owes much to Nunn and Qian (2011).

Finally, our work is related to several empirical papers in economic geography where
large shocks are often identified as natural experiments. Davis and Weinstein (2002) is the
seminal contribution, but recent important contributions include Redding and Sturm (2008)
and Bleakley and Lin (2012). Much of this work is focussed on the question of permanence.
What causes permanence in the location of production and what does this tell us about
competing theories of economic geography. Our work fits well in this tradition although it
has some unique elements. It documents strong permanence within a given regime (both
the Organic and Mineral economy periods); an extreme reshuffling across regimes; and ties
a change in the locational fundamentals of energy supply to this shift. Although locational
fundamentals have often been cited as important or even pivotal, their very permanence
has made it difficult to assess their importance empirically. An energy transition however
provides exactly this needed variation.

The rest of the paper is organized as follows. In section two we introduce our Energy-
Economy model where energy demand from firms interacts with energy supply to determine the rate of exploitation for a single renewable energy source. Sections 3 contains our Malthusian application and a discussion of data and hypothesis testing is contained in section 4. A short conclusion ends the paper. All proofs and lengthy calculations are relegated to the appendix.

2 The Energy-Economy Model

We assume consumption and production activities are located at an economic core while potential energy sources are distributed in the surrounding space. The economy’s core contains all of its production and consumption units but is zero dimensional. The exploitation zones where energy sources can be found are two dimensional planes allowing us to employ definitions of area, distance, and density. Distance is meant to capture any and all costs incurred when incremental amounts of energy are exploited.

2.1 Tastes, Technology, Endowments

There are $L$ identical consumers each endowed with one unit of labor. Their utility is defined over the consumption, $C$, of a final output good $Y$ which provides energy services. Consumers’ income comes from providing labor services and from reaping resource rents. Since nothing important hinges on consumer numbers we will model a representative consumer with labor endowment $L$. Utility of our consumer is strictly increasing and strictly concave in consumption:

$$u = u(C), \quad u' > 0, \quad u'' < 0$$  \hfill (1)

Final output $Y$, is produced *inter alia* by a set of goods intermediate in the supply chain between energy exploitation and final consumption. Naturally we will refer to them as intermediate goods. We take them to represent a set of capital goods and consumer durables tailored to a specific energy source. These intermediate goods are of course the backbone of any real world energy system since they are the means by which energy services are delivered to firms and consumers in the economy.

While any one firm or individual may employ only a subset of these intermediates to provide energy services, aggregate output in the economy would be a function of all available intermediates. One simple, tractable, and common way to capture the reliance of final output on the set of intermediates is to adopt a constant elasticity Dixit-Stiglitz specification
for final goods production.

\[ Y \equiv \left[ \sum_{i=1}^{n} \frac{x_{i}^{\sigma}}{x_{i}^{\sigma-1}} \right]^{\frac{\sigma}{\sigma-1}}, \quad \sigma > 1 \tag{2} \]

where each of the \( x_i \) are differentiated intermediate goods produced by a single firm. Together these intermediates provide the service flow represented by the final output \( Y \).

It is important to our analysis that the set of available intermediates is endogenously determined. To do so we follow standard practice and assume \( \sigma > 1 \) to allow for their production by monopolistically competitive firms. The number of firms is then determined by a free entry condition. Each of these firms incur both fixed and variable costs to produce any intermediate. Let \( a_i \) be the activity level of firm \( i \), then the production of a typical intermediate at level \( x_i \) requires \( a_i \) units of activity. That is:

\[ a_i = \alpha + \beta x_i \quad \forall i \tag{3} \]

where \( \alpha \) represents the set up cost to produce a typical intermediate while \( \beta \) represents its constant marginal cost of production in terms of activity.

Activity is the employment of labor in conjunction with power. Specifically, the activity of each firm, \( a_i \), requires both a flow of energy and labor given by the following constant returns function:

\[ a_i = f(l_i, W_i) \quad \forall i \tag{4} \]

We take (4) to be a Cobb-Douglas function with a labor share in the value of output given by \( z \). Constant returns is innocuous given increasing returns are already assumed in the mapping from activity to intermediate output; the constant shares assumption is useful in identifying specific regions of the parameter space to characterize results.

Combining these assumptions we see that final energy services to consumers (and producers) are provided by a set of differentiated processes that use both conventional inputs (labor) and energy. In the abstract, these processes are energy converters that provide the services of heat, light and power; and any energy converter that transforms energy into services with the help of conventional inputs is broadly consistent with our formulation (a toaster, flashlight, and laser printer all fit the bill). In the concrete, we have ignored the particularities of these converters by assuming all intermediate goods are symmetric substitutes; and importantly, we have adopted a formulation where the greater is the number of ways energy services can be delivered, the higher is the system’s overall productivity in translating energy and inputs into services.

\[ ^{3}\text{The service interpretation is well known and well used in the international trade literature. See for example, Markusen (1990).} \]
2.2 The Demand for Energy

The demand for energy comes from two sources. Energy is used in the production of intermediates, as well, energy is used directly in transportation. It proves useful to treat energy used in transportation as an input use problem, and we leave it to our discussion of energy supply in the next section.

As is well known, the mark up rule and zero profit condition of the monopolistic competition framework allow us to neatly solve for the level of output and activity per firm. Shephard’s Lemma then returns the implied factor demands for labor and energy per firm. Together these are simply:

\[ x_i^* = \frac{\alpha}{\beta} (\sigma - 1) \]
\[ a_i^* = \alpha \sigma \]
\[ l_i^* = \alpha \sigma \frac{\partial c(w, p^W)}{\partial w} \]
\[ W_i^* = \alpha \sigma \frac{\partial c(w, p^W)}{\partial p^W} \]

Since all firms are identical, total energy demanded (for production purposes) is just \( n \) times firms specific demand \( W_i^* \).

The number of firms, \( n \), is endogenous and depends on the overall market size as indexed by \( L \). To solve for the number of firms use the firm specific demands for labor and note full employment requires \( \sum_{i=1}^{n} l_i^* = L \). By substituting for \( l_i^* \) we can solve for the number of firms:

\[ n = \frac{L}{l_i} = \frac{L}{a_i^* z c(w, p^W)/w} = \left( \frac{1 - z}{z} \right)^{1-z} \left( \frac{L}{a_i^*} \right) \left( \frac{w}{p^W} \right)^{1-z} \]

because \( w l_i = z a_i^* c(w, p^W) \). We can now add up across firms to find aggregate energy demand, \( W^D \) (Watts demanded) as:

\[ W^D = n W_i = L \left( \frac{1 - z}{z} \right) \left( \frac{w}{p^W} \right) \]

The properties of demand follow directly from our assumptions. Constant returns in activity implies an increase in market size or \( L \), raises demand proportionately. Constant factor shares implies a unitary price elasticity of demand.

2.3 The Supply of Energy

The collection and transportation of resources, like everything else, requires power. The net power supplied to the core is equal to total power collected in the exploitation zone, \( W \), minus the power needed for transportation. In obvious notation, \( W^S = W - W^T \).
2.3.1 Power Used in Transportation

The area exploited to find and collect energy is related to the power demands of the core measured in Watts, [W], and the power density, $\Delta$, measured in Watts per squared meter, [W/m$^2$], of the particular energy source exploited. If the flow of needed power is $W$, then the area exploited must equal:

$$EX = W/\Delta$$  \hspace{1cm} (8)

where $EX$ is measured in $m^2$.

If the energy resources are distributed uniformly over land, then minimum cost search implies energy will be collected from a circular area with the economy’s core at its center. The average carrying distance, $ACD$, that needs to be covered to bring the energy resources inside the circle to the core is given by:

$$ACD = \int_0^R r[2r/R^2]dr = \frac{2}{3}R$$  \hspace{1cm} (9)

where $R$ is the radius of the circle delimiting the exploitation zone. Since Area is $A = \pi R^2$ and $EX$ represents the area of exploitation, we have $ACD = (2/3)(EX/\pi)^{1/2} = (2/3\sqrt{\pi})(W/\Delta)^{1/2}$ using (8).

The mass of the energy resources transported is equal to the area under exploitation times the (uniform) physical density of the fuel over this area, $M = EX \times d[kg]$, where $d$ is in units of kg/m$^2$. Force is equal to mass, $M$, times acceleration $g[m/s^2]$; as any mass moved horizontally must overcome the force of gravity as mediated by friction in transport. The coefficient of friction is given by $\mu$ and it has no units. Work is in turn force times distance, where distance is given by (9). All this implies:

$$Work = \mu g M \times ACD = [Newton.meters] = [Joules]$$  \hspace{1cm} (10)

$$Work = \frac{2}{3} \mu gd \left(\frac{W}{\pi \Delta}\right)^{3/2}$$

where we have used (8) and (9). $Work$ is then measured in Joules. This work is done per unit time since power is flow as is the flow of labor services and the flow of useful output. If we measure time in seconds, then the flow of $Work$ measured in Joules per second is now power requirements measured in Watts.$^4$ So the total energy cost of delivering the flow of

$^4$Expending a Joule of energy per second means you are delivering power of one Watt.
power $W$ to the core is given by $TC(W)$:

$$W^T = \frac{2}{3} \mu g d \left( \frac{W}{\pi \Delta} \right)^{3/2} \equiv TC(W)$$

(11)

Note $TC(W)$ is a strictly convex function of energy output, $W$, where $TC(0) = TC'(0) = 0$.

With these preliminaries in place we can now consider the economics of energy supply.

### 2.3.2 Energy Production

We assume a sole energy supplier chooses the area of exploitation to maximize flow profits taking prices as parametric.\(^5\) The energy supplier’s optimization problem is to choose total power to maximize profits:

$$\max_W \Pi = p^W W - (p^W + p^C) TC(W)$$

(12)

where $W$ is total power embodied in the energy resources exploited, $p^C$ is the price of the final output, and $TC(W)$ is defined above.

There are three assumptions reflected in (12). The first concerns how we have dealt with the issue of net power. By construction, the firm chooses total power to maximize profits, but supplies to the market only that power net of use in transportation. It in effect sells $W$ watts of power at $p^W$ but buys back from itself $TC(W)$ watts to pay for transportation back to the core.\(^6\) By making power requirements proportional to total carry we have stayed true to the physics of transport highlighted in the Only Energy Model of Moreno-Cruz and Taylor (2013).

The second assumption is that transportation uses both energy and the final output good as inputs. The assumption provides a link between the efficiency of providing energy services and the cost of energy exploitation. We mentioned previously how lower energy prices created entry into intermediate good production and raised productivity; now we see that this higher productivity (created, for example, by an energy price decline) will imply lower exploitation costs and greater energy supply. This assumption closes the positive feedback loop in the energy system.

The final assumption is the limited substitutability between energy and final output in transport. Every unit of energy used in transportation must be matched with a unit of final output.

---

\(^5\)Assuming a single producer is innocuous, since the natural alternative of assuming atomistic energy suppliers produces identical results. Assuming the supplier maximizes flow profits is innocuous when energy density is not affected by the degree of energy exploitation.

\(^6\)Electric transmission pays a transport cost via transmission losses; natural gas pipelines run their turbines on natural gas; diesel fuel runs diesel fuel tanker trucks, etc.
output to provide transportation. This assumption reflects a critical modeling decision. It ensures the energy costs of energy exploitation are bounded above zero. To ensure the energy input needed per Watt of power collected can never approach zero even in the limit, we need a production function for transport with an elasticity of substitution between energy and conventional inputs of less than one. Within this class of functions, the Leontief specification delivers the simplest results. One way to think about this assumption is that it respects the physical reality that friction cannot be eliminated - even in the limit - regardless of the scale of conventional inputs applied in transportation.

With these assumptions in place, the firm’s maximization problem has a simple solution:

\[
\left( \frac{p^W}{p^W + p^C} \right) = TC'(W)
\]

(13)

Let \( s \) be equal to the share of energy costs in extracting and transporting resources to the core, then \( s = p^W / [p^W + p^C] \) and the solution to our firm’s problem becomes \( s = TC'(W) \).

Since \( TC'' > 0 \), we can invert (13) to write total power as \( W^* = \psi(s) \) and power supplied as \( W^S \equiv \psi(s) - TC(\psi(s)) = g(\psi(s)) \). It is now simple to show:

\[
\frac{dW^S}{ds} = [1 - s]\psi' > 0
\]

(14)

where use has been made of (13). Not surprisingly, the supply of power to the core is increasing in the relative price of energy to a unit of final output.

Since energy demand is a function of relative factor prices, we need to put supply in similar terms. To do so recall that final good production is a constant returns activity, and hence active production requires its price equal unit cost. This implies

\[
p^C = c(p_1, p_2, ..., p_n) = \left[ \sum_{i=1}^{n} p_i^{1-\sigma} \right]^{\frac{1}{1-\sigma}}
\]

(15)

Since all firms are identical, we can substitute and rearrange to find:

\[
p^C = \left[ \sum_{i=1}^{n} p_i^{1-\sigma} \right]^{\frac{1}{1-\sigma}} = \frac{\beta \sigma}{\sigma - 1} c(w, p^W)n^{\frac{1}{1-\sigma}}
\]

(16)

The unit cost of energy services reflects not only the cost of labor and energy, but also the set of intermediate goods used in producing the service flow. An increase in the set of intermediate goods tailored to a energy source raises overall productivity and therefore lowers the costs of energy services delivered.
To simplify further use the linear homogeneity of $c(w, p^W)$ and (6) to substitute for the number of intermediate goods to find:

\[
\frac{p^W}{p^C} = \frac{s}{1-s} = z \left( \frac{1-z}{z} \right)^{\frac{\sigma(1-z)}{\sigma-1}} \left( \frac{\sigma-1}{\beta \sigma} \right) \left( \frac{L}{\alpha \sigma} \right)^{\frac{1}{\sigma-1}} \left( \frac{p^W}{w} \right)^{\frac{z(\sigma-1)}{\sigma-1}}
\]

(17)

And hence we can now link energy supply to relative factor prices to find:

\[
\frac{dW^S}{d(p^W/w)} = \psi'[1-s] \frac{ds}{d(p^W/w)} > 0 \iff \sigma > 1/z
\]

(18)

Recall $\psi' > 0$. The definition of $s$ and inspection of (17) signs the derivative.

Somewhat surprisingly, we need to impose a parameter restriction in order to ensure the supply curve for power is positively sloped. The reason is simple: cheap energy creates more energy converters; more converters raises productivity; higher productivity lowers the costs of energy exploitation, and this in turn makes energy cheap. With these positive feedbacks built into the model, we have unbounded increasing returns at the aggregate level.\(^7\) Since unbounded increasing returns is an unattractive feature of any model, in Moreno-Cruz and Taylor (2012) we bound increasing returns with a simple parameter restriction. Bounding increasing returns ensures energy supply is always upward sloping and combining this with our energy demand produces a unique general equilibrium.\(^8\)

But even with bounded IRS the model admits two cases: an Almost Neoclassical case where increasing returns are weak and there are no incentives for agglomeration; and an Increasing Returns case where increasing returns are stronger and incentives for agglomeration exist. While these two cases deliver similar results in a variety of settings, in the interest of space we will focus entirely on the Increasing returns case which obtains when $\sigma < 3/[1-z]$. We assume this condition holds throughout.

\subsection*{2.3.3 Characteristics of Supply}

Two characteristics of energy supply feature prominently in much of what follows. The first is simply that even though the potential supply of energy is infinite, it is effectively

\(^7\)By using (5) we can write final output solely as a function of the number of firms producing intermediate goods and parameters. Going further we can use (6) and (7) to write the number of intermediates as a constant returns function of labor input and watts delivered to the core. Putting these results together we find the economy’s aggregate production function relating final energy service output to the economy’s (inelastically supplied) endowment of labor and the (endogenously determined) supply of power: $Y = BL^a W^b$ with $a + b > 1$ where $a = \frac{z \sigma}{\sigma-1}$, $b = \frac{[1-z] \sigma}{\sigma-1}$, $B > 0$. To bound increasing returns we assume both $a$ and $b$ are less than one.

\(^8\)We prove this in the appendix.
bounded by a physical limit. To see why recall \( W^S \equiv \psi(s) - TC(\psi(s)) = g(\psi(s)) \), and substitute for the particulars of the model to obtain total power \( W^* = \psi(s) \), and supplied power \( W^S = g(\psi(s)) \) in explicit form as follows:

\[
W^S = \left[ 1 - \frac{2}{3} s \right] W^* \quad W^* = \frac{\pi s^2 \Delta^3}{[\mu gd]^2} \quad R^* = \left[ \frac{s \Delta}{\mu gd} \right]
\]

where we note \( s \equiv s(p^W/w, L) \) from (17). The reason for the upper bound on power supply can now be explained straightforwardly. As energy prices rise, more power is naturally supplied but this is sourced from more distant (read more difficult) sources. More distant sources carry with them larger transport costs and this pushes the share of energy costs \( s \) in transport costs towards their maximum of one. When \( s \) approaches one, marginal energy resources take as much energy to collect and transport as they provide - implying their contribution to net power supply is zero. At this point, it doesn’t matter how valuable another unit of energy is - net supply becomes perfectly inelastic.

The implication for us is that even though a city or urban area is in some sense small in terms of the surrounding forest or farmland - our core here is after all a point in a limitless plane of energy resources - physics alone will at some point limit the exploitation zone that can fuel this city. Even a small urban center in a huge country may reach its peak draw on neighboring energy resources long before they are exhausted.

The second feature of note appears, at first blush, to be quite specific to our circular formulation. It is the fact that total power is a cubic function of power density. As we will show subsequently, it arises from the spatial context and not the particular form of our exploitation zone. The intuition is instructive. Consider the thought exercise of increasing the power density of neighboring resources while holding energy prices constant. If we do so, but leave the area of exploitation fixed; then supplied power will rise proportionately with power density; i.e. appear with power 1 in an expression like (19) because \( W = \Delta EX \). But a higher power density also implies the marginal cost of exploitation falls. With lower marginal costs, exploitation rises and the extensive margin of exploitation moves outwards. Since area is proportional to the square of radius, total power rises with the square of power density. Adding up over both margins implies supply is a cubic of power density.

This scaling law implies that even small differences across resources in their power density can create large differences in equilibrium prices, quantities, etc. And hence if we thought that moving from an energy resource like wood with an energy content of 15 MJ/kg to even poor quality coal at 30 MJ/kg represents a small change in economic circumstances we could be very wrong.
2.3.4 Energy Prices

While there are several energy prices in the model we will focus on the relative factor price of power, $p^W/w$ since this is relevant to firm’s decisions concerning energy use; and the relative price of energy services to the wage, $p^C/w$ since this is relevant to the welfare of residents in any location.

Given our earlier discussion on the limits to supply, it should not come as a surprise that sufficiently large increases in market size, or $L$ will drive energy prices upwards. Less obvious though is that energy prices may also fall over some range.\footnote{In fact, $s$ is always rising in $L$.} The reason is simply that increases in market size create efficiency gains in the energy delivery system and these gains can produce declining prices over some range. These gains benefit both consumers - who see their purchasing power in terms of energy services rise - and producers - who see their costs of transport fall. The strength of increasing returns in the model determines whether these efficiency gains are large enough to offset the direct costs of sourcing more expensive marginal energy resources. Under the Increasing Returns case we are considering, we find:

**Proposition 1** An increase in market size at first lowers the relative price of energy services to the wage but eventually leads to an increase.

Under slightly stronger conditions ($\sigma < 3$) the relative factor price of energy also falls and then rises with market size as shown in Figure 1.

2.3.5 Consumption and Output per person

Given the last proposition it may not be surprising that the relationship between market size and consumption per person are determined by similar forces.

**Proposition 2** Consumption per capita is single peaked in market size.

Proof: see Appendix.

**Proposition 3** Consumption per capita rises with power density. The peak level of Consumption per capita occurs at a larger market size when the power density of available resources is greater.

Proof See Appendix.

We depict these results in Figure 1(a) where we have assumed power densities $\Delta'' > \Delta' > \Delta$. Proposition 2 confirms what we may have already suspected: the interaction of IRS
in the energy system with diminishing returns created by our spatial structure produces a
singular peak in consumption per capita. Proposition 3 then tells us this peak is not only
higher, but at a larger market size when resources are more power dense. These results
are of course critical in that they establish a link between the power density of available
resources and incentives for agglomeration.

The cost curves shown in Figure 1(b) reflect exactly the same forces and the two panels
are linked. The trough of energy service price occurs at a smaller market size than does
the peak of consumption per capita; and within Figure 1(b) there is another relationship.
Firm’s face a rising cost of energy before consumers do. The trough in terms of the real
factor price for energy, always occurs before the trough of the relative price of energy services
to the wage. The reason is simply that consumer’s benefit from the fact that large markets
provide many ways to deliver energy services, and these complementarities allow them to
insulate themselves from rising power prices at least for some time. Intermediate goods
producers who hire power and labor directly have no such ability. One last result is that
the gap between these two relative prices grows with market size as shown in the figure.

![Consumption and Prices](image)

(a) Consumption Percapita  
(b) Relative Prices

Figure 1: Consumption and Prices

The implication of these results for our work is simply that consumer’s may be more
than willing to move to a city with steeply rising power costs, since the market size bene-
fits provided by a larger agglomeration at least temper and can even reverse any negative
implications on their own consumption per capita.

2.3.6 Roads, rivers and canals.

We have thus far assumed a very diffuse transportation system: every energy producer
makes their own way to the core. This may be a reasonable representation for farmers
bringing their produce to markets, but it excludes the possibility that the nature of energy
resources themselves may affect the set of transportation options. For example, rivers, roads and canals were all important features of the energy transportation system in 1800 when resources were primarily shipped in their solid and bulky form; just as power lines, oil pipelines, and LNG terminals are important features today when resources are shipped as liquids or electricity. What all of these options have in common is that they represent low friction and presumably low cost methods of transporting energy to markets. And while some transportation options are given by nature (rivers) or provided for other means (roads), large scale transport of energy resources requires at least the maintenance or improvement of these facilities to provide low cost transport. This was surely the case during the time period we study because roads, rivers, and ports were improved; while canals and railroads were new innovations driven by the advent of the coal economy. We now investigate how rivers, roads, and their improvements affect the motives for agglomeration.

To start we assume any given energy supplier can take energy directly into the city or deviate to take advantage of a road or river nearby. Rivers and roads help to reduce the amount of work used in transportation, increasing the amount of energy delivered to the city. To capture this in our analysis we allow for the coefficient of friction of the river or road to differ from the coefficient of friction of land by a fraction \( \rho < 1 \). That is, while the coefficient of friction of land is equal to \( \mu \), a road’s coefficient of friction is \( \rho \mu \) in both directions whereas when traveling with the current a river’s is also \( \rho \mu \) but against it \( \mu / \rho \). By this assumption, river transport is only useful when you are an energy producer upstream; whereas road transport reduces frictions in two directions and not one.

We assume the river or road is a straight line that crosses the core of the city and expands indefinitely. The location of a supplier relative to the city is described by two terms: \( \iota \), the distance from the city and \( \theta \) the angle between the segment formed by the city and the supplier and the river. Suppliers decide how to travel to the core, that is what fraction of the trip is done by river or road and which distance is done by land. The road allows for energy producers farther from the core to profitably bring energy to it, but there are still limits to how far suppliers are willing to take their energy. Panel 2(a) shows how the exploitation zone is altered by the presence of a road. In Panel 2(a) the angle \( \bar{\theta} = \arccos(\rho) \) shows the limiting angle that in combination with the limiting radius \( \bar{\iota} \) determine the profitable zone. Suppliers inside this zone will bring their energy to the city, suppliers outside this zone will not. Comparing Panel 2(a) to Panel 2(b) shows the gains introduced by a road relative to the case in which there is no low cost transportation alternative. Moreover, as the transportation infrastructure becomes more competitive, that is \( \rho \) decreases, the critical angle and the limiting radius both increase. The case of a river is similar, but only half the gains are accrued because the river is one-directional.
We can now calculate the energy supplied to the city when a river is available:\(^{10}\)

\[
W^S = \frac{\Delta^3}{(\mu gd)^2} s^2 \left( 1 - \frac{2}{3} s \right) g(\rho) 
\]

\[
g(\rho) = \pi - \bar{\theta} + \int_{0}^{\bar{\theta}} \left( (1 - \rho^2)^{1/2} \sin \theta + \rho \cos \theta \right)^{-2} d\theta \geq 0
\]

where the function \(g(\rho)\) is positive and monotonic and approaches infinity as \(\rho\) goes to zero. Setting \(\rho = 1\) means the river or road offers no advantage in terms of transportation. This implies \(g(\rho) = \pi\) since then \(\bar{\theta} = 0\) and equation (20) reduces to equation (19).

Inspection of equation (20) reveals that the role of improved transportation is identical to being granted a more dense resource base in terms of energy supplied to the core. River or roads multiply by \(g(\rho)\) the power density of available resources to an extent determined by its capacity for reducing transport costs as reflected in \(\rho\). An important feature of this result is that the effect of the improved transport, in terms of energy supply, is increasing in the density of available resources; therefore, improved transport is most powerful for very dense resources, and least powerful for low density resources.

This is a very important result because it extends all our earlier results on power density to results concerning the existence of rivers or roads for transport. For example, we know that a unique general equilibrium still exists; the existence of a river or road (modeled as a movement from \(\rho = 1\) to some \(\rho < 1\)) shifts the supply curve for energy outwards lowering the relative price of energy while raising the equilibrium energy supplied, final output and the number of intermediate goods in the energy system; access to a river or improved road

\(^{10}\)The derivation is similar to the one we show in Moreno-Cruz and Taylor (2013). Readers are directed there for more details.
transportation raises both consumption and output per capita at all population levels; and
it raises the market size at which output and consumption per capita peak.

2.3.7 Endogenous Infrastructure

We now examine the incentives for improving and maintaining transportation options. For example, roads need improvement and maintenance, rivers need dredging, locks need cleaning, etc. In many cases these investments are undertaken by central authorities or undertaken by private corporations under contract. During our time period, parliament passed individual bills contracting out for both road improvements (turnpike bills), and river improvements (navigation acts). To see how these options may work, we assume a single agent chooses how much to improve the infrastructure in order to maximize the flow of profits from energy sales. The problem for this central planner is:

$$\max_{\rho} \Pi^N = p^W W^S(\rho) - (p^W + p^C) h(\rho)$$

(22)

where $W^S(\rho)$ is given in (20) and $(p^W + p^C) h(\rho)$ are the flow costs of investing in infrastructure. For simplicity we have assumed one unit of the composite is used alongside each unit of energy spent in building and maintaining infrastructure. We assume these costs are an increasing function of the reduction in friction so that $h'(\rho) < 0$ and $h''(\rho) > 0$. The first order condition that maximizes profits requires

$$\frac{\Delta^3 s^3}{(\mu gd)^2} \left(1 - \frac{2}{3}s\right) g'(\rho) = h'(\rho)$$

This equation implicitly defines $\rho$ as a function of power density, $\Delta$, and the energy share $s$. Using this implicit function we can ask how power density affects the optimal investment in infrastructure and the economic implications of this investment. The next proposition summarizes these results:

**Proposition 4** Higher power density implies greater infrastructure investment, lower energy prices and greater supply.

**Proof:** See Appendix.

3 The Malthusian Application

We now develop a Malthusian model for population growth to examine the interaction of power density and population density. The Malthusian mechanism is simple, well under-
stood, and arguably appropriate given our empirical application covers years well before any demographic transition; indeed our data comes from a time (1086-1841) and a place (England) where Malthus himself observed the forces he identified at work.\textsuperscript{11} We start by examining an Isolated State, but since villages, hamlets, and cities do not exist in isolation, we also extend our analysis to allow for migration both within and across regions.

### 3.1 The Isolated State

We adopt a very standard Malthusian specification where population growth responds to increases in consumption per capita, but amend it to allow for a higher death rate arising from crowding. There is a baseline birth rate, $\eta_0$, and baseline death rate, $\delta_0$, with $\eta_0 - \delta_0 < 0$ representing the net rate of population reduction when consumption per capita is zero. This baseline population growth is then adjusted by assuming births rise proportionately with consumption per capita, $\eta_1 c$ and deaths fall proportionately with consumption per capita $\delta_1 c$, where $c = C/L$. During the Malthusian era, death rates in cities were known to be much higher given their poor sanitation and crowding. We incorporate this into our population dynamics by assuming death rates due to crowding rise proportionately with population density, $\delta_2 L$. As a result, the birth and death rate functions are given by:

\begin{align}
\eta(c) &= \eta_0 + \eta_1 c \\
\delta(c, L) &= \delta_0 - \delta_1 c + \delta_2 L
\end{align}

Hence the population responds according to:\textsuperscript{12}

\begin{equation}
\dot{L} = L[\eta_0 - \delta_0] + [\eta_1 + \delta_1]c - \delta_2 L
\end{equation}

In the standard Malthusian set up $\delta_2 = 0$, and subsistence consumption is a constant independent of city size. Once we allow city size to affect death rates, consumption must rise with city size to hold population growth at zero. Define the Zero Population Growth or

\textsuperscript{11}There is an ongoing debate over the usefulness of the Malthusian model as a means to understand population growth in England over this time period. While there are several issues at dispute, one basic problem is that almost any population history can be rationalized by the Malthusian model if we allow for shifts in the preventive check over time. Since these shifts are usually motivated by forces outside the model, they are clearly difficult to evaluate empirically. Consequently, while Clark (2007) presents an unabashed Malthusian view of the world prior to 1800, Allen (2008) disputes much of this evidence and argues the application is simplistic and often problematic. A more favorable recent evaluation, using cross country data to the year 1500, is presented in Ashraf and Galor (2011).

\textsuperscript{12}An implicit assumption in this analysis is that everyone lives in some form of agglomeration, be it beside a manor, village, town, or city. Adding a fixed purely rural population is easy to accomplish but adds little to our results.
$Z-line$ as the combination of consumption per person and population consistent with zero population growth. Setting (24) to zero the $Z-line$ can be written as:

$$c_Z(L) = \frac{\delta_0 - \eta_0}{\delta_1 + \eta_1} + \frac{\delta_2}{\delta_1 + \eta_1}L$$

where $c_Z(L)$ is the per capita consumption needed to maintain zero population growth in a city with population size $L$.

Since we have already characterized how consumption per person changes with market size we now combine these two constructs to identify the possible steady states of our system in Figure 3. The three $c(L; \Delta)$ lines drawn in the figure show how consumption per capita varies with $L$ in regions that differ in the power density $\Delta$ of available resources. If available resources are not very power dense, then the $c(L; \Delta)$ curve lies everywhere below the $Z-line$ and a vacant landscape results. When resources do offer sufficient power density, then two possible city sizes emerge: small cities (the leftmost intersections) and large cities (the rightmost intersections). It is easy to see that with Malthusian population dynamics, the small city steady states are not stable whereas the large city steady states are stable. This implies that small cities, when they exist, are fragile constructs. A small negative perturbation leads to their disappearance as the region converges to the (stable) vacant land steady state; a small positive perturbation starts the process of transformation of the small city into a large city. The figure shows two regions where agglomerations can survive. It is straightforward to see that the intersection of the $Z$-Line with the consumption per-capita curve $c(\Delta_H, L)$ occurs at a higher level of consumption and for larger population sizes than the intersection with the $c(\Delta_L, L)$ curve. Hence, higher power density implies higher consumption per-capita and larger population sizes. This discussion leads us to the core result of our paper:

**Proposition 5** *The density-creates-density hypothesis:* There exists a critical level of power density, $\Delta^{\text{crit}} > 0$; if power density is below this critical value, $\Delta < \Delta^{\text{crit}}$ then the only steady state has $L = 0$; (iii) if $\Delta > \Delta^{\text{crit}}$ then there are three steady states: steady state zero has zero population, $L_0 = 0$, has a vacant landscape, and is stable; steady state one has a positive population, $L_1 > 0$, is unstable, but has a relatively low population; and steady state three has a positive population, is stable, and has a larger population $L_2 > L_1 > 0$. The population level in steady state three is increasing in power density $\Delta$.
3.2 Within and Across County Migration

Thus far we have assumed our cities are truly Isolated States: their own population growth determines market size. Since there is no inter-city migration, and no simultaneous small or large city pair feeding and sustaining each other, we have only a very partial understanding of how the distribution of power densities across space may influence the distribution of cities and their size. The ability of the population to migrate changed greatly over our 800 year sample period, and therefore it is important to understand how migration interacts with the distribution of power densities to determine city sizes.

We introduce migration by constructing a simple three point distribution linking power density to city sizes and migration. This example highlights many of the features of the mapping from power density to city size, and has the benefit of being transparent.

3.2.1 A 3 County Example

To understand how flows of migration affects city size, consider Figure 4. We have drawn three Isolated States which differ in the power density of their neighboring resources. By assumption, $C$ has the highest power density and $A$ the lowest. Also shown in Figure 4 is the Malthusian or $Z$-line. The intersection of this line with our three consumption per capita curves generates the associated stable steady states labelled $A$, $B$ and $C$.

Now consider the migration decision. Although agents may be inclined to move to the
most productive region since consumption per capita is highest there, so too are death rates. To see why recall that death rates equal birth rates along the Z-line, and birth rates rise with consumption per person. This implies that all upward movements along the Z-line correspond to situations where agents face a higher death rate. The simplest way to interpret the Malthusian model is that everyone in the city (man, woman, child, king or queen) faces the same instantaneous probability of death given by $\delta(c, L)$. If we accept this fiction and assume a zero discount rate for simplicity, then the expected lifetime consumption at time $\tau$ for any individual living in city $i$ is given by:

$$e = \int_0^\infty c_i \exp^{-\{t-\tau\} \delta(c, L)} \, dt \quad (26)$$

$$e = c_i / \delta_i \quad (27)$$

where the last equality holds in steady state. This construct now allows us to compare expected lifetime consumption in the various locations.

Start with city $B$, which we will refer to as the middle city. At $B$, there is an associated consumption per person and population size; and using (26) we can calculate the expected lifetime consumption per capita any resident of $B$ would obtain, denote it $e_B$. It proves useful to construct a line representing all combinations of $\{c, L\}$ that yield the same expected
lifetime consumption per capita any resident would obtain at the steady state $B$:  

$$c_E(e_B, L) = \left[ \frac{e_B}{1 + \delta_1 e_B} \right] [\delta_0 + \delta_2 L]$$

(28)

which we will refer to as the $e_B$-line. Several properties of $e_B$-line follow directly. Holding expected lifetime consumption constant, the $e_B$-line is just a linear relationship between population size and consumption per capita. Points above the $e_B$-line correspond to combinations of consumption and population sizes yielding higher expected lifetime consumption than $e_B$; points below correspond to lower expected lifetime consumption than $e_B$. Finally, although it is not obvious, the $e_B$-line going through steady state $B$ is flatter than the associated $Z$-line going through the same steady state.

It is immediate then that points labelled $A'$ and $C'$ in Figure 4 offer the same expected lifetime consumption per capita as at $B$. Point $A'$ is however above the $Z$-line: its consumption level exceeds that necessary to generate zero population growth and, in the absence of migration, its population level would necessarily rise from $L_{A'}$. Indeed - without migration - its population would increase until the city reached the no-migration steady state at $A$. An alternate possibility is that this constant excess flow of population could migrate instead to city $C$ thereby holding its current population at $L_{A'}$ and its current expected lifetime consumption equal to that offered at $B$. Now consider city $C$. Residents in this city would have the same expected lifetime consumption as those at $B$, if it was to grow and reach point $C'$. But of course point $C'$ is below the $Z$-line, and hence consumption per capita at this level falls short of that necessary to generate zero population growth. Without further adjustments, its population would decline from the level $L_{C'}$ towards its no-migration steady state at $C$. The now obvious alternative is for city $C$ to remain at $C'$ and receive a constant flow of population from city $A$. If the flows from city $A$ to $C$ balance we have identified a steady state of the system with migration.\(^{13}\)

In theory, Figure 4 shows how a given distribution of power densities across previously Isolated States maps into a steady state distribution of city sizes with active migration. We have assumed here that cities differ in the power density of their available resources as they would if they came from different counties. This interpretation is however overly restrictive. Cities within the same county may differ in power densities if some have access to low friction alternatives like rivers or roads while other cities in their county do not. Therefore, our three cities could be in one county, two counties, or three counties.

The migration steady state has several important features. First, migrants flow from

\(^{13}\)Since consumption per capita schedules are continuous in power density we can always construct a steady state of this type; we are not claiming this steady state is unique.
small cities to big cities. Big cities offer higher consumption per capita but higher death rates. Migration does not equalize real wages across the population centers, but it does equalize expected lifetime consumption. Second, migration widens the distribution of city sizes. The biggest city got bigger and the smallest city got smaller. Third, for any city or region sending migrants we find a rise in productivity as consumption (and output) per person rises. In addition overall productivity may also rise - migrants move to a region with higher average productivity and lessen the Malthusian pressures that keep productivity low in their sending regions. Fourth, migration is associated with an increase in the birth rate in sending regions but depresses it in the large city receiving it.

4 The Population History of England 1086-1841

The population history of England has been studied by legions of academics with demographers, economists, historians and medievalists alike producing literally hundreds of contributions dissecting and debating the period from the Norman conquest to the very first years of Queen Victoria’s reign. Demographers in particular have spent decades carefully collecting and analyzing data from this period, and here we mine this very deep vein of scholarship to present for discussion a view of this period’s history through the lens of our theory.

We start with a descriptive sketch of the period using simple statistics and data representations. Much of what we say will not surprise experts of the period as we employ standard sources, and present no new controversial claims. Indeed our descriptions of the data owes much to the earlier work of Wrigely (2010) and Allen (2009) both of whom examine the role energy played in shaping English history over much of this period. Following our sketch we construct a causal argument linking the growth of London, the transition to coal, and flows of migrant workers to changes in the distribution of economic activity across time and space.

4.1 A Sketch of 800 Years

The population history over our entire sample period is shown in Figure 5. Panel 5(a) presents population estimates for the years 1086, 1290, 1377, 1600, 1801 and 1841. These six population figures, together with their county specific population levels, are the core

\[14\] Several excellent book length treatments by Sir Tony Wrigley present a fascinating picture of English demography and economics over this time period. Especially relevant are Wrigley (2010) that focusses on energy issues; Wrigley (1969) and Wrigley (1978) are also valuable resources. Allen (2009) presents an engaging and rigorous examination of the forces leading to the Industrial Revolution focussing on energy sources, induced innovation and the role of international trade.
data we employ. Panel 5(b) converts raw county population figures into county specific population density measures, and then smooths them with a kernel density estimator. The panel shows a plot of these kernel density estimates for our 39 English (ancient) counties for our six snap-shot years. To make comparisons easy we present these figures in logarithms, rotate the densities by 90° and order them chronologically.

As shown England starts with a 1086 population of approximately 1.7 million and ends with a 1841 population of a little under 15 million. This growth however represents an anything-but-constant average population growth rate of 0.3 percent per annum. In fact this period is dominated by two cycles of growth and subsequent plateau, punctuated by the worst epidemic in recorded human history. The earliest period from 1086 to 1290, although affected by the Norman conquest and the resulting political upheavals it created, was a relatively prosperous time for the population. Population growth during the thirteenth century was robust and the population of close to 4.5 million citizens in 1290 may have represented the country’s peak population given its almost exclusive reliance on low density organic energy sources at this time. There is no mention of coal use in the Domesday book of 1086, although there were small shipments of coal into London during the 13th century. We view the 1086 to 1290 period as one where England’s economy was organic, labor mobility was low, and the Malthusian forces we have highlighted played an important role in population growth. The 1290 population figure may have represented something close to a Malthusian steady state for England.15

The Black Death struck England in the spring of 1348 and swept across the country with great intensity for the next two years. Although records are far from exact, the population plunged with those infected suffering a 70% mortality rate. Three additional outbreaks occurred during the next fifty years, but this first outbreak created the greatest and most widespread loss of life. Not surprisingly, estimates for the aggregate population in 1377 are well below the 1290 peak. During the next three centuries the population recovered slowly only surpassing its pre-plague levels by perhaps 1600. Although it is not apparent from Panel 5(b), the 1600 to 1841 period did not exhibit constant population growth.16 The period started with a minor resurgence of the plague in 1625 and a major one in 1665; the great fire of London followed in 1666 and the Restoration in 1688. For almost the entire century England is actively at war with either Scotland, France, or Spain. These and other yet-to-be understood forces produced relatively slow aggregate population growth until the 1730s. Thereafter population growth rose sharply.

By the end of our sample period, England was a key colonial player, the world’s leading

15 This seems to be the view of many economic historians. See for example Broadberry et al. (2011).
16 See for example the discussion in Wrigley (1969) Chapter 3 and especially Figure 3.3.
power, a prolific trading nation, the world’s largest coal producer, and home to the world’s largest and most cosmopolitan city. Coal use rose significantly over this early modern period starting with the introduction of a significant coal trade to London in the fourteenth century. At this time, most use was for domestic heating although coal also found uses in a range of industries from lime and salt production to smelting. New mines opened in western England, Scotland, and Wales and as the economic incentives for adopting coal use grew technical barriers were overcome in a whole host of industries where coal replaced charcoal or wood. During this period coal moved from being exclusively a source of domestic and industrial heat to a provider of mechanical power via the Newcomen and Watt steam engines. In 1841 the wondrous Great Western Railway from London to Bristol went into operation,
ushering in a rapid period of railroad expansion throughout the country. London was by this time the world’s largest city with a population well over one million people.

4.2 History through the Lens of Theory

Our theory makes several predictions linking the distribution of population over space to the distribution of energy resources. To make any headway at all in mapping one to the other, much simplification is required. For example, we assume that in addition to the forces identified in our theory exogenous (and unmodelled) technological progress drives long run economy wide changes in population levels; we ignore almost completely the important role international trade may have played in these events; and we focus almost exclusively on our model’s steady state outcomes. In doing so we limit the discussion to elements which are novel to our theory.

Without doing too much violence to the historical record we divide our almost 800 year sample period into two. We refer to the first period as the Organic Economy period, since during this time the vast majority of energy used by households and industry was provided by renewable sources. We date it from 1086 to 1377. It starts twenty years after the Norman conquest and ends shortly after the first and most severe ravages of the Black Death. We refer to the second period as the Mineral Economy period and date it from 1377 to 1841. This second period starts after the Black Death and runs until the very start of the Victorian era in 1841.

4.3 The Organic Economy

Consider the initial two snapshots \{1086,1290\} shown in Panel 5(a). These are years untouched by the plague. Our interpretation of this data is straightforward. With little migration across counties and diffuse natural blessings across England, the geographic distribution of the population should likewise be diffuse; and the upward shift in the distribution over time reflects nothing more than the workings of slow but steady technological progress relaxing the bonds of an organic economy. To make these connections precise we first divide our 39 counties into three groups, and then place them in a Malthusian steady state as shown in Figure 6 below. As shown one group is labelled \( M \), and this group contains just the county of Middlesex (which contains London). A second group is the set of 25 counties without coal deposits. We label these \( NC \) for no coal. The final group is the set of lucky 13 counties with coal, labelled \( C \). The corresponding steady states occur at the intersections \( O_M, O_{NC}, \) and \( O_C \) where the county specific consumption schedules intersect the zero population growth Z-line as shown in Figure 6.
In constructing Figure 6 we have made several assumptions. The first is the ordering of groups. In 1086, Middlesex county - which contains London - is the most dense county in England. This is of course long before London became a hub for international trade, a recipient of coal shipments or even the capital of England. London’s location on the Thames (which was cleverly chosen by the Romans) did however give it a far wider exploitation zone than inland cities with similar natural environments. The Thames is a tidal river meaning ships can both enter and leave with the current, while London’s location is far enough inland to protect from attack and storms but close enough to benefit from tidal flows and a deep channel. During much, if not all, of this time period London was sourcing wood resources from the Thames valley far up river and coast-wise as well.\textsuperscript{17} According to Proposition 4, a low friction alternative such as the Thames magnifies the power density of surrounding resources and should raise the density of population centers as well.\textsuperscript{18} Whatever advantage the Thames and the existing Roman Road system granted to London, it was limited in an organic economy setting.\textsuperscript{18}

\textsuperscript{17}See, for example, Figure 1, p. 459 from Galloway et al. (1996). The figure, although for a somewhat later period, shows London’s zone extends up the Thames valley past Henley, and includes almost all of Middlesex, and parts of Surrey, Berkshire, Oxfordshire, Buckinghamshire, Essex, Kent, and Hertfordshire.

\textsuperscript{18}See Hilbert (1977) for a discussion of early London. Since Londinium was in fact a Roman invention, it was also the beneficiary of an ancient roman road system. Consulting maps will show that London is the hub of the Roman system with "almost all roads leading to London".
The second assumption is that the relevant (i.e. stable) steady states occur on the downward sloping segments of the consumption per capita schedules. This reflects our belief, and that of significant scholarship in this area, that in some sense England was already full at this time and diminishing returns were beginning to be felt in agriculture. Our placement of the steady state past the peaks, implies further population growth would lower real wages however measured. The third assumption is zero or low labor mobility. Since the institution of serfdom and demesne farming were still very much in play, labor mobility was low. This assumption appears uncontroversial.

While our steady state configuration may capture important cross sectional aspects of these two distributions, it says nothing about the shift in the distribution over time. Since the benefits of nature are fixed over time, our theory suggests that the population distributions in 1086 and 1299 should be almost identical once we allow for at most trend growth tied to technological progress. While this may be true, the population level and its geographic distribution in 1086 also reflects the impact of the Norman conquest in 1066. William the Conqueror put down a series of revolts in northern England and these losses are apparent in the lower tail of the 1086 population distribution. Once we correct for mean differences across the years, the two (recentered) distributions are quite similar although this data will reject the null that the 1086 and 1290 (recentered) distributions are in fact identical. However if we exclude ANY one of the four most Northerly counties bordering Scotland - Northumberland, Cumberland, Westmoreland or Durham - we cannot reject the null hypothesis that the (recentered) distributions of population across the remaining counties is the same in both 1086 and 1290 at any reasonable level of significance. For example excluding Northumberland produces a p-value of 0.147 using the Kolmogorov-Smirnov test for equality of distributions. If we look deeper at either the raw correlations across county population densities or their rank correlations across these two periods, we again find strong evidence of permanence. The raw correlation, again without Northumberland, is 0.8; the rank correlation is 0.7.\footnote{Further evidence that these wasted counties were in the process of catching up can be provided by a simple convergence regression or plot of subsequent population growth rates on levels in 1086. The differential equation governing population growth in the Isolated State case can easily be log-linearized and then manipulated to generate a convergence style regression with log population growth rates related to a common to counties constant and county specific initial log population densities. A plot of growth rates versus levels along these lines is provided in our earlier discussion paper. See Moreno-Cruz and Taylor (2012).}

The happy state of affairs circa 1300 soon changed. Early in the century a great famine struck Europe precipitated by repeated crop failures, and in 1348 the Black Death hits and lowers populations everywhere by almost one half. In terms of our theory this moves the system immediately to the positions we show in Figure 6 as the post plague outcomes at
\(P_M, P_{NC}, \text{ and } P_C\). Real wages rise dramatically as a result of this decline, and soon after we observe the 1377 populations in our data.\(^{20}\) The plague lowered the national population tremendously, but as the density for 1377 shows the incidence of death from the plague was more severe in already dense counties. Despite this complication, if we use the same K-S distributional test as previously we cannot reject the hypothesis that the (re-entered) distributions for 1290 and 1377 are in fact the same; as is 1086 and 1377 when we exclude any one northern county from the test. Similarly, using the full set of counties, we find the raw and rank correlations across the last two years of the Organic period are also quite high at .81 and .83.

The Organic period writ large features a nation recovering after the conquest with wasted counties catching up while others grew slowly towards their limits. There was little change in the geographic distribution of population which is as it should be when migration is limited and key energy resources are constant over time. Even the most lethal epidemic in human history appears to have had little impact on the geographic distribution of the population.

4.3.1 The Trigger

Although the plague may not have altered the geographic distribution of population very much, it still plays two very important methodological roles in our research design. Its first role is conventional: the plague provides a very large shock to the existing distribution of population. This shock may have loosened the grip that previous centers of agglomeration had on the economy allowing the post-plague period to differ from the pre-plague period. Using a major shock to explore the strength of agglomeration forces was first pioneered in seminal work by Davis and Weinstein (2002). This methodology has subsequently found wide use within the empirical literature in the new economic geography. While the plague was certainly a shock of the first order, it also plays a second and more important role in our analysis. Its second role arises from the very significant labor shortage caused by the depopulation. This shortage strengthened the hand of the laboring classes and, together with institutional changes already underway, led to the breakdown of the villeinage system over the next century and a half. Demesne land becomes smaller, more land is contracted out, and laborers gain far more freedom to move. While historians continue to debate whether the Black Death was the key driver in these changes or whether it was merely one of the forces at play, this debate matters little here because the facts are well established.\(^{21}\)

\(^{20}\)In our theory it is possible for real wages to fall if the population decline is severe enough to move us far past the peak of the consumption per person schedule and close to zero output in a location.

\(^{21}\)See the very nice article by Hatcher (1994) where he reinvigorates this debate by arguing for a pivotal role of the Black Death.
The breakdown of serfdom, the change in land practices, and the growth in mobility over this period are well documented. These institutional changes now imply laborers are far more able to migrate within and across counties in search of better opportunities. In terms of our theory, this greater freedom in the centuries following the Black Death means that England does not return to anything like its earlier 1290 or 1377 configuration of Isolated States; it returns instead to a world where labor is free to move both within and across counties.

4.4 The Mineral based Economy

Consider the next three snapshot years \{1600, 1800, 1841\} in Panel 5(a). These years show a quite different pattern emerges as we enter what historians often refer to as the the Early Modern Era. They show a radically changing and changed England. By the year 1600 two features stand out. The first is what appears to be a narrowing of the population distribution across a majority of the counties. The second is the run away growth of Middlesex county. By 1600 Middlesex county had a density almost 9 times greater than its next competitor (Surrey); while the same calculation in 1290 shows the ratio of Middlesex to the next competitor (Suffolk) is less than 2. In 1841, a similar feature appears, but now Middlesex is only 6 times as dense as its closest competitor. Density is however now an astounding 2169 people per square kilometer. In total the distributions shift upwards over time; develop a long right tail; and, have their mass shift left.

These aggregate features are largely driven by the growth of London. London rose from 50,000 in 1520 to perhaps 200,000 in 1600. Growth the following century was also rapid. By 1700 London had over half a million inhabitants; it represented over 10% of the English population; and was by then Europe’s largest city. Growth slowed closer to national averages over the next century but London was 900,000 by 1800. By 1841 it was well over a million and the world’s largest city.

We examine these events with the help of Figure 7 (where we have cut out part of the diagram to focus our attention). The figure shows the Organic Economy configuration previously discussed plus a set of outcomes \{M_M, M_{NC}, M_C\} associated with a free migration steady state. The ability to migrate has several implications. First, real wages are higher than otherwise in the sending groups and lower than otherwise in the receiving group. We see at least the beginnings of convergence in real wages across the country. Second, average productivity in the sending groups is higher because Malthusian constraints are relaxed; it is lower than otherwise in the receiving group. Overall productivity nation wide may well be higher since migrants move to the most productive location. Third, we have drawn the figure with both the $C$ and $NC$ group as net suppliers of migrants, while Middlesex
is a net demander. That there was large scale migration to London during this period is not at issue; the evidence for real wage convergence is however more mixed. But the concentration of population in London has an interesting implication: the migration steady state we identify by \( \{M_M, M_{NC}, M_C\} \) cannot be right; in fact, as we argue below its very construction contains a contradiction.

![Figure 7: The Organic Economy with Migration](image)

**4.4.1 The Transition**

If the new configuration representing the migration steady state at \( M_M, M_{NC}, \) and \( M_C \) was right, it would have to be consistent with a London (and hence Middlesex) population of 200,000 in 1600. But 200,000 Londoners is far beyond the estimates we have of London's ability to feed and fuel its populace in an Organic economy setting.\(^{22}\) Work by economic historians mapping out both London's fuel and food exploitation zones, suggests a London population of 80,000 already strained natural resources; a population well over 100,000 would have brought sharp consequences, but by the end of the 16th century London housed 200,000 people. And therein lies the rub: we are forced to conclude that the concentration of energy demand in London, created by migration, brought about its energy transition to coal in the 16th century. In short, London's growth brought about large economic incentives to solve the relatively small but numerous technical problems limiting the use of coal in domestic heating, and in other industrial processes where the fumes or chemical deposition arising

from open pit burning of coal fouled the production process.

London’s incipient energy crisis was solved by the transition to coal in a significant set of uses. While this view is not universally shared, the transition to a denser energy (mix) would now shift the consumption per person schedule for Middlesex upwards (not shown), raise expected lifetime consumption in the county, and perhaps even accelerate migration to the capital. London shakes off the bonds of the Organic Economy and enters the Mineral Economy period. And while the Thames is a constant advantage to London, its impact is all the greater when the energy we source is more power dense. London, already blessed by geography, is now blessed by the geographic location of vast north east coal reserves which are easily shipped coast wise and up the Thames.

4.4.2 The Treatment

What happens next is shown in Figure 8. Ignore for the moment the dashed line. As shown there is active migration from both NC and C to M given the location of the $Z^{-}$ line also drawn. This initial situation would represent sometime in the 16th century when London has already moved significantly to coal. At this point, the same migration flows that drove the energy transition centered in London, now provide a mechanism for the diffusion of this new information regarding coal uses throughout the country. These novel ways to use coal are perhaps interesting to residents of NC counties, but not very useful since the overland costs of transporting coal were then prohibitive. In C counties, this information is both useful and productive. Cheap coal and information from London combine to produce a much faster and relatively cheap transition to a Mineral based economy in all C counties.

In terms of our theory, the consumption per person schedule for all the C counties shifts upwards and leapfrogs that of the NC counties. This knock-on effect of London’s transition is shown by the dashed curve labelled $C^*$ in the figure.

The resulting change in economic opportunities in C counties then redistributes population towards the C counties by both raising their natural growth rates and lowering any migration they sent to London. The remaining NC counties do not benefit immediately from coal, and fall behind in relative terms. These two very different responses create a tremendous reordering of the rank in population densities across counties while allowing population densities themselves to approach unheard of levels. By the 18th century growth of London and Middlesex slows as it was more difficult to attract migrants from the now industrializing coal areas. And soon after we obtain the observations for 1800. By 1841 little has changed.

Several facts support this view. First, migration to London was a first order phenomenon. Wrigley (1967) for example constructs a simple numerical example using estimates of death
and excess birth rates of the day to show how London would need to harvest virtually all the excess births in the country to grow at its spectacular pace. His estimates led to his oft quoted claim that one in six Englishmen or women of this period spent some of their life in London. Second, we know that gross migration flows must have been much larger than net. Estimates of net flows are in the tens of thousands per year. Third, a significant portion of the migrant population - apprentices - came for the express purpose of learning new skills and transferring them back to their home counties; that is, they came to facilitate technology transfer. Apprentices were a very important part of the London population representing somewhere between 10 and 20% of the London (i.e. up to 40,000 individuals in London of 1600). Therefore, not only was their significant migration to London, it was two way migration, and a significant fraction of these migrants were just the right sort to bring new technical solutions developed in London home for productive use in their counties. Fourth, a variety of statistical evidence is at least consistent with the view that there was reordering of the rank in population density. The most dramatic of this evidence is shown in Figure 9 where we plot the rank order of population density in the Malthusian steady state year of 1290 versus the population distribution in the Mineral Based economy in 1841.

If the rank order of counties did not change across these periods, then the counties most densely populated in 1290 would also be those most densely populated in 1841: all points would lie on a 45 degree line (Middlesex does lie on this line with rank number 1). If there was a strong relationship over time, the data should show a strongly positive relationship.

---

23See Field (2010) for reference to primary sources supporting these numbers.
As shown, there is very little - if any - relationship across these periods.

Figure 9: Population Rank: 1290 vs. 1841

The implication of these exercises is clear and immediate - the distribution of population densities created by the advent of coal and the beginnings of a mineral based economy in the early modern era had little to do with the densities created by the forces of the organic economy determining outcomes in 1290. A list of major cities and their population circa 1086 or 1290 with a comparison in 1841 will show the same result, as has the comparison of county densities. With few exceptions, the population redistributed itself over this period from traditional centers of population to new coal rich areas.\textsuperscript{24}

\subsection*{4.4.3 Research Design}

The perfect experiment evaluating our density-creates-density hypothesis would of course be the random assignment of energy resources with different power densities across otherwise identical counties who had little to no interaction via either trade or migration. Random assignment of resources with different power densities in a world of Isolated states would cause an initial supply shock to treated units, lower energy prices, raise consumption per capita and spur infrastructure investment. Over the longer term we would find population density was caused by differences in the power density of available resources. While the real world is not this neat, our research design strives to replicate at least part of this fiction, and control for those aspects we cannot manipulate.

\textsuperscript{24}The key additional exception will be Lancashire that both grew very fast and very early from the innovations introduced in the cloth industry. Lancashire does have coal, but its growth will be shown to be only minimally determined by its deposits.
Our discussion of the history suggests we adopt a simple difference in difference research design where changes in the population density in both $NC$ and $C$ counties is linked to the advent of information on how to use coal. The Organic economy period before 1600 is pre-treatment \{1086, 1290, 1377\}; the Mineral economy period is post treatment \{1600, 1800, 1841\}. The impact of coal is identified by exploiting our sample’s specific cross sectional ($NC$ vs. $C$) and over time (Organic period vs. Mineral period) variation.

### 4.4.4 Timing

As we mentioned earlier, our view of when the transition occurred is not universally shared. For example, while Hatcher (1993) claims that by 1600 coal was London’s predominant fuel, both Allen (2009) and Fouquet (2008) suggest somewhat later dates. Our argument for an earlier date relies on the logic of the argument presented above; the fact that coal exports to London per person were higher circa 1600 than they were in any future year; a myriad of other small pieces of information suggesting London merchants, home owners, and even cloth dyers may have shifted to coal by this time; and empirical evidence we will subsequently present that accords well with this timing. But we are not suggesting London moved entirely to coal by this time, and to some extent our difference with earlier authors is a matter of degree and not substance. We date the time information becomes freely available about coal’s uses as 1600 onwards, but we examine whether our results are sensitive to this assumption. In addition, we provide placebo tests by moving the treatment window backwards in time to investigate sensitivities. These investigations allow the data to speak regarding the correct length and placement of the treatment window.

### 4.4.5 Reverse Causation

We have excluded Middlesex for exactly this reason, and in fact rely on a causal relationship between the growing size of Middlesex and the transition to coal. It is however possible that population growth in other counties in turn sped or caused their movement to coal as well. While this is certainly possible $C$ counties were on average smaller and less dense than their $NC$ counterparts pre-treatment, so if large populations drove a movement to coal (rather than cheap local access) we should find no significant treatment effect. Coal was used in almost all counties of England over much of this time period, but it was typically employed in a relatively few applications, was generally viewed as inferior, and unless local demand made firewood very dear - coal would see little use. Smithing may be the only example of an industrial demand where coal was superior to wood. But we have already seen that over the period in question, almost all of the growth in England was centered in London. This
observation makes it very difficult to sustain a belief that a localized shortage of wood in coal counties drove their transitions.

While it is true that several counties surrounding London were likely influenced by London’s growth and also had access to coal, they are also NC counties. Whatever coal inspired growth they achieved would only dampen the treatment effect we estimate.

4.4.6 Other Causes

It is possible that C counties were rapidly growing over the treatment period for other reasons - perhaps trade related - and those counties primarily benefitting from trade also happen to have coal. In an effort to mitigate this concern we will control for the existence of large and small ports and two different measures of existing road networks in our empirical work. To capture their time varying effects, we will interact these characteristics with a full set of time dummies to allow their impact to vary over time. As a consequence if C counties grew because they also contained ports booming from international trade, we will, in theory, already be accounting for this variation. Our measure of roads and ports will be constructed using pre-treatment measures (Roman roads, roads at the time of Domesday, and historically important ports).

What is critical to our research design is that most of the technical problems surrounding the use of coal in large industrial and household uses were solved first in London; and that there exists a realistic way for these small incremental innovations to spread fairly uniformly throughout England. We have argued that the constant flow of migrants into London - those seeking work, bright lights, or education via explicit apprenticeships were the perfect transmission mechanism for the many small innovations and local learning in London, but we will also present evidence in support of this view as well.

With regard to London’s innovativeness we are also on firm ground as several authors have argued for similar effects. Allen (2010) for example argues that the incremental innovations needed in home building to take advantage of coal benefitted from local learning and experimentation in London and then spread throughout the country.

4.5 Empirical Implementation

4.6 Data

We have collected data from several different sources. Our key population data is taken from recent estimates by Broadberry et al. (2011). We employ this data plus measures of county areas to generate a series for population density across counties and time. In addition we
have exploited information on coal mines, coal output and coal trade from both volume I and II of Hatcher’s classic treatise on the UK coal industry (1993). Data on international trade was collected from a variety of sources, as was information on the road and transport network in both Roman times and the date of Domesday. In some cases, recourse to historic maps was necessary to locate county towns, the extent and quality of coal fields, etc. A complete list of these resources and means to construct our data is presented in the data appendix.

4.7 Baseline Specification

Our empirical implementation is fairly straightforward. We seek to compare the growth in population density over time in counties with and without coal deposits. While we treated all C coal counties as identical and all NC no coal counties as identical, they are in fact very different. Since counties differ in size, climate, topography and political organization we include county fixed effects in all specifications; similarly a varying climate (both natural and political), the Black Death, and technological progress call for us to include unrestricted time dummies in all specifications. In addition, we introduce variation in the treatment intensity across counties that is proportional to the size of exposed coal deposits. This measure of coal availability is taken from maps of the day and it is determined entirely by geology. Considering the extent of exposed deposits provides us with a further source of identification: within-coal-county, county variation. For comparison purposes we also provide estimates from a simple binary treatment reflecting the presence or absence of coal deposits in the county. Surprisingly, all fifteen coal fields in England were already active prior our treatment period and therefore issues of initial discovery are largely moot.

Our Baseline specification follows very closely that adopted in Nunn and Quian (2012). It is described by:

\[
dens_{it} = \beta \text{coal\_indicator}_i \cdot I_t^{\text{post}} + \sum_{j=1086}^{1841} X'_i I_{jt} \nu_j + \sum_{j=1086}^{1841} \gamma_j I_{jt} + \sum_{c=1}^{38} \rho_c I_{ci} + \epsilon_{it}
\]

where \(dens_{it}\) is the log of population density in county \(i\) at time \(t\); the \text{coal\_indicator} is either a binary indicator or the log of the percent coal field area in a county; \(I_t^{\text{post}}\) is an indicator variable taking the value one in 1600, 1800, and 1841; \(I_{jt}\) is an indicator variable taking a value one when \(j = t\) and is zero otherwise; \(\gamma_j\) are the time fixed effects; and \(\rho_c\) the county fixed effects. The coefficient of interest is \(\beta\), which is the estimated impact of the percent of coal area (or the binary indicator) on population density. We present the results from our baseline estimates in Table 1 below.
Table 1: Baseline Estimates With Percent Coal

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variable: Ln Population Density</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent Coal 1600-1841</td>
<td>0.273***</td>
<td>0.273***</td>
<td>0.279***</td>
<td>0.277***</td>
<td>0.273***</td>
</tr>
<tr>
<td></td>
<td>(0.0521)</td>
<td>(0.0535)</td>
<td>(0.0519)</td>
<td>(0.0522)</td>
<td>(0.0540)</td>
</tr>
<tr>
<td>County Fixed-effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Time Fixed-effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Strong-Port</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Weak-Port</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Strong-Road</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Weak-Road</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>228</td>
<td>228</td>
<td>228</td>
<td>228</td>
<td>228</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.856</td>
<td>0.857</td>
<td>0.862</td>
<td>0.868</td>
<td>0.898</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As we move across the table additional controls are added to the basic specification in column 1 that has only time and county fixed effects. Columns (2) first adds controls for strong (or important) ports in a county while weak (or less important) ports are added in (3). Both are interacted with time dummies to allow these to vary as growth or trade changes. Similarly columns (4) and (5) control for the density of the existing Roman road network (strong roads) and the entire set of roads available at the time of Domesday (weak roads). It is apparent that in all specifications, the treatment effect is positive and significant. Its magnitude varies only slightly across specifications and is about 0.27. This estimate implies a doubling of the area covered by coal in the county (an increase of 100%) would raise population density by approximately 27% - not a small amount.

Since the coverage of coal by counties varies quite significantly from over 50% coverage to less than 5% this estimate implies the movement to coal implied quite different things to different countries. One way to see this is to ask how much of the growth that occurred in a county from 1377 to 1841, was due to coal? To calculate this figure we first find the overall population growth for each county and then subtract for each county the contribution created by coal using its county specific measure of coal coverage area and the coefficient estimate from our full specification in column (5). The result from this exercise is the
amount of growth that would have occurred in the absence of coal. Comparing this growth to the estimate of the growth caused by access to coal, give us the share of a county’s growth that we can attribute to coal. We do this for every county and graph the result in Figure 10 below. The estimates range from a low of 11% of growth for Somerset to over 60% for Nottinghamshire! For four counties the growth from coal is less than 25%, but for five others it is above 50%. The remaining group fall somewhere in between, but in all cases these shares indicate coal played a very significant role in growth for coal counties.

![Figure 10: Share of Growth Created by Coal](image)

Our results demonstrate that coal was important to the growth of coal counties, but was it key to the redistribution of economic activity over this period? To answer this question we present three population distributions in Figure 11 below. The top two distributions plot log population densities in 1290 and 1841 and we distinguish the 13 lucky coal counties with dark bars. Looking just at these two panels reveals a strong correlation between having coal deposits, and moving up in the overall distribution. 11 of the 13 coal counties lie below the median in 1290; 10 of 13 lie above it in 1841; and 3 of the 5 most dense counties in 1841 are coal counties! In the last panel we have calculated a counterfactual distribution by eliminating the growth effects we attribute to coal (using table 1, column (5) estimates). This counterfactual distribution gives a smaller role to coal deposits: five coal counties now move to below median growth, all others shift down the distribution giving up their top spots; and only one in five of the most dense are now coal counties. Therefore, coal was important to all coal counties, but only pivotal for some.
Figure 11: Coal and the Distribution of Population
4.8 Sensitivity Tests

In Table 2 below we investigate the implications of a simple binary indicator for a coal county. In this case the impact of coal is identified from just across county and across time variation. One reason we may be interested in this measure is if we think the coal coverage ratio, however constructed, is contaminated because it reflects perhaps local demand for coal. If true, this would amount to reverse causation. While we do not believe this is the case, the estimates in Table 2 are likewise consistent in their finding of a statistical significant positive impact on population density. As well, the economic significance is also large. The estimates from column (5) indicate growth on the order of .65 log points; indicating a coal county would all else equal be 1.9 or almost twice as dense than otherwise. Perhaps not surprisingly, similar results are obtained when we use the area of coal fields in the county.

Table 2: Baseline Estimates With Binary Indicator

<table>
<thead>
<tr>
<th></th>
<th>Dependent Variable: Ln Population Density</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Binary Coal 1600-1841</td>
<td>0.583***</td>
</tr>
<tr>
<td></td>
<td>(0.127)</td>
</tr>
<tr>
<td>County Fixed-effect</td>
<td>Yes</td>
</tr>
<tr>
<td>Time Fixed-effect</td>
<td>Yes</td>
</tr>
<tr>
<td>Strong-Port</td>
<td>No</td>
</tr>
<tr>
<td>Weak-Port</td>
<td>No</td>
</tr>
<tr>
<td>Strong-Road</td>
<td>No</td>
</tr>
<tr>
<td>Weak-Road</td>
<td>No</td>
</tr>
<tr>
<td>Observations</td>
<td>228</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.844</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In Table 3 we investigate the timing of the transition. Since it is not entirely clear that our timing of events is correct, we now investigate the implications of a more flexible estimation. In this case, the post treatment dummy is replaced by a full set of time dummies. By doing we obtain an estimate for the interaction of coal county and time period by period. Since the initial period is the base year the value of these interactions is not important but
their time pattern is relevant. Table 3 contains estimations that allow for a completely flexible treatment structure. Importantly, the coefficients start relative small, fall slightly in 1377 but then start to rise sharply with over a 50% gain during this critical interval. Further upward growth is seen in 1800 and 1841. A visual representation of these estimates together with their 95% confidence interval is shown in Figure 12 below. The figure suggests a distinctly rising pattern of impact after 1600 which is of course our treatment period.

![Figure 12: 95% Confidence interval flexible estimates](image)

In Table 4 we go further and take our three year treatment window and investigate alternative placements of this window first starting in the initial 1086 period, and ending in our treatment window that starts in 1600. The initial window is just the opposite of our treatment period and hence must generate results identical to the true window after sign changes and account for constants. As we move rightwards the negative coefficients disappear by 1377 and both middle periods show no significant effect. Moving the window back into our chosen post treatment period reestablishes our results. The final output from this table is one more flexible estimation but this time within the treatment window. As shown each of the years makes a significant contribution to the treatment again suggesting our post treatment dating of 1600 was appropriate.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent Coal 1290</td>
<td>0.220*</td>
<td>0.218*</td>
<td>0.215*</td>
<td>0.215*</td>
<td>0.207*</td>
</tr>
<tr>
<td></td>
<td>(0.0965)</td>
<td>(0.101)</td>
<td>(0.0968)</td>
<td>(0.0915)</td>
<td>(0.0919)</td>
</tr>
<tr>
<td>Percent Coal 1377</td>
<td>0.193**</td>
<td>0.192**</td>
<td>0.194*</td>
<td>0.194**</td>
<td>0.171**</td>
</tr>
<tr>
<td></td>
<td>(0.0612)</td>
<td>(0.0661)</td>
<td>(0.0734)</td>
<td>(0.0662)</td>
<td>(0.0578)</td>
</tr>
<tr>
<td>Percent Coal 1600</td>
<td>0.306***</td>
<td>0.306***</td>
<td>0.309***</td>
<td>0.308***</td>
<td>0.270***</td>
</tr>
<tr>
<td></td>
<td>(0.0650)</td>
<td>(0.0691)</td>
<td>(0.0767)</td>
<td>(0.0705)</td>
<td>(0.0654)</td>
</tr>
<tr>
<td>Percent Coal 1801</td>
<td>0.447***</td>
<td>0.445***</td>
<td>0.450***</td>
<td>0.449***</td>
<td>0.443***</td>
</tr>
<tr>
<td></td>
<td>(0.0656)</td>
<td>(0.0704)</td>
<td>(0.0769)</td>
<td>(0.0696)</td>
<td>(0.0593)</td>
</tr>
<tr>
<td>Percent Coal 1841</td>
<td>0.480***</td>
<td>0.479***</td>
<td>0.486***</td>
<td>0.485***</td>
<td>0.484***</td>
</tr>
<tr>
<td></td>
<td>(0.0727)</td>
<td>(0.0768)</td>
<td>(0.0799)</td>
<td>(0.0733)</td>
<td>(0.0620)</td>
</tr>
<tr>
<td>County Fixed-effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Time Fixed-effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Strong-Port</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Weak-Port</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Strong-Road</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Weak-Road</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>228</td>
<td>228</td>
<td>228</td>
<td>228</td>
<td>228</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.872</td>
<td>0.873</td>
<td>0.878</td>
<td>0.884</td>
<td>0.912</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Table 4: Treatment Estimates With Percent Coal

<table>
<thead>
<tr>
<th>Percent Coal</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1290-1600</td>
<td>-0.231*** (0.0628)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1377-1801</td>
<td>0.0645 (0.0801)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1600-1841</td>
<td>0.273*** (0.0540)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1600</td>
<td>0.145* (0.0543)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1801</td>
<td>0.317*** (0.0470)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1841</td>
<td>0.358*** (0.0503)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>County Fixed-effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Time Fixed-effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Strong-Port</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Weak-Port</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Strong-Road</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Weak-Road</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>228</td>
<td>228</td>
<td>228</td>
<td>228</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.888</td>
<td>0.866</td>
<td>0.898</td>
<td>0.905</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
5 Conclusions

In this paper we developed a simple spatial model where energy sources differed in their power density, where the collection and exploitation of energy was costly, and where improvements in the energy system were complementary to energy exploitation. Within this context we showed how the concept of power density could play a major role in determining the size and density of urban agglomerations. The power density of an energy source determines the peak level of consumption per capita; roads and rivers effectively magnify the power density of available resources; and differences in power density across environments could create large differences in populations. We termed this last result the density-creates- density hypothesis.

To investigate this hypothesis we needed a specific mechanism to sort population across competing locations; a period when the energy system was relatively simple; and a place where data on energy sources and agglomeration was available. The natural solution was to use a Malthusian mechanism to investigate the population history of English counties over the 800 year period from the Norman invasion to the reign of Victoria. The deep scholarship of this period provided to us data, many insights, and documentation of the first energy transition the world has traversed.

With help from our theory, we were able to develop a narrative account of this period consistent with many accounts of previous researchers. To move beyond this narrative we used our theory to suggest a very natural research design allowing us to investigate the causal mechanism we proposed. Our empirical results show that the movement to coal in our first energy transition was an important part of the massive reshuffling of the economic landscape. In five of the Lucky 13 coal counties, access to coal accounts for over 40% of their population growth; for all but one of this group access to coal provided over 20% of their population growth. For several counties, coal was absolutely critical; for all others it was an important driver of growth.

6 References


A Proofs

**Existence:** Demand for energy is decreasing in $p^W/w$, it approaches infinity when $p^W/w$ approaches zero and it approaches zero when $p^W/w$ approaches infinity. From equations (14) and (18) energy supply is increasing in $p^W/w$ as long as $\sigma > 1/z$. To prove existence it is enough to show the supply function is zero when $p^W/w$ is zero. From inspection of (17) we know that $p^W/p^C$ is zero when $p^W/w$ is zero, which implies $s = 0$. From (13) this implies $TC'(W)$ is zero which by equation (11) implies energy supply is also zero.

**Proposition 1:** Total differentiating the demand and supply equations and rearranging we find

$$\hat{W}_S = \frac{6(1-s)}{3-2s} \hat{s} + 3\hat{\Delta} \quad \text{and} \quad \hat{W}_D = \hat{L} - \left(\frac{p^W}{w}\right)$$

(A.1)

where a variable with a hat on top means percentage change. By the definition of $s$ and equation (17) we find

$$\hat{s} = (1-s) \left(\frac{p^W}{p^C}\right)$$

(A.2)

and

$$\frac{p^W}{p^C} = \left(\frac{z\sigma - 1}{\sigma - 1}\right) \left(\frac{p^W}{w}\right) + \left(\frac{1}{\sigma - 1}\right) \hat{L}$$

(A.3)

Equalizing equations in A.1, setting $\hat{\Delta} = 0$ and rearranging terms we find

$$\frac{\left(\frac{p^W}{w}\right)}{\hat{L}} = 1 - \frac{6(1-s)^2}{3-2s} \frac{1}{\sigma - 1} - \frac{6(1-s)^2}{3-2s} \frac{1}{\sigma - 1} + 1$$

(A.4)

Therefore, the relative price of power to the wage raises with market size if

$$\sigma - 1 > \frac{6(1-s)^2}{3-2s}$$

(A.5)

The righthand side of the inequality is bound between 0 and 2 (because $s$ can only take values between 0 and 1). When $\sigma \geq 3/[1-z]$ the lefthand side of the inequality is larger than 3 so that the inequality always holds.

Next, take equation (17) and multiply both sides by $w/p^W$ and find:

$$\frac{\hat{p}^C}{w} = \frac{\sigma(1-z)}{\sigma - 1} \left(\frac{p^W}{w}\right) - \left(\frac{1}{\sigma - 1}\right) \hat{L}$$

(A.6)
Replacing (A.4) in (A.6) we find

\[
\frac{\mu^C}{w} = \frac{\sigma(1-z) - 1 - 6(1-s)^2}{3-2s} \frac{1}{\sigma-1} + \frac{6(1-s)^2 z\sigma - 1}{(3-2s) (\sigma-1)} \frac{1}{\sigma-1}
\]  

(A.7)

Therefore, the relative price of manufactured goods increases with the size of the economy if and only if

\[
\sigma(1-z) - 1 > \frac{6(1-s)^2}{3-2s}  
\]

(A.8)

As before, the righthand side in the inequality (A.8) is bounded between 0 and 2. When \( \sigma > 3/(1-z) \) we can see the lefthand side is always greater than 2 which ensures \( \rho^C/w \) increases monotonically with the size of the economy. When \( \sigma < 3/(1-z) \) the lefthand side of (A.8) is less than 2. Thus, for \( s \) close to zero the condition is violated and an increase in market size lowers the relative price of energy services to the wage. However, as the size of the market increases, \( s \) also increases. To see this replace equations (A.3) and (A.4) back in equation (A.2) to show:

\[
\frac{\sigma}{L} = \frac{6(1-s)^2 z\sigma - 1}{(3-2s) (\sigma-1)} + \frac{z\sigma}{\sigma-1} > 0
\]  

(A.9)

Hence, as the size of the market increases, \( s \) approaches 1 and the condition in (A.8) eventually holds.

**Proposition 2:** Consumption is defined as the residual of output that remains after transporting the energy to the city: 

\[ C = Y(L, W^S) - TC(W^*, \Delta) \]  

where \( Y(L, W^S) \) is aggregate output, \( TC(W, \Delta) \) in given in (11) and \( W^S = W^* - TC(W^*) \). We are interested in characterizing consumption per capita \( C/L \) as a function of \( L \) and \( \Delta \). Total differentiation of consumption per-capita \( C/L \) yields:

\[
\frac{d(C/L)}{dL} = \frac{dC}{dL} \frac{1}{L} - \frac{C}{L^2}
\]  

(A.10)

After some manipulation we show in Moreno-Cruz and Taylor (2012) we find

\[
\frac{d(C/L)}{dL} = \left( \frac{1-z}{\sigma-1} \frac{6(1-s)^2 z\sigma - 1}{(3-2s) (\sigma-1)} + \frac{z\sigma}{\sigma-1} \right) \frac{Y}{L^2} - \frac{C}{L^2}
\]  

(A.11)

Now we are ready to show that consumption per capita peaks when \( \sigma < 3/(1-z) \). To show that consumption per capita is single peaked, \( \frac{d(C/L)}{dL} \) must be positive for small values of \( L \), negative for large values of \( L \) and the sign of \( \frac{d(C/L)}{dL} \) can change only once. First we
find the critical value of $L$ where the slope is zero by setting (A.11) equal to zero:

\[
\frac{1 - z}{\sigma - 1} \frac{6(1 - s)^2 z \sigma}{16(1 - s)^2 (z \sigma - 1) + (3 - 2s)(\sigma - 1)} + \frac{z \sigma}{\sigma - 1} = \frac{C}{Y} \tag{A.12}
\]

Taking the derivative with respect to $L$ and replacing equation (A.9) we can show that $dLHS/dL < 0$. From the definition of $s$ we know that as the size of the market increases, the price of energy services to the wage also increase and $s$ approaches 1. This implies that the LHS approaches $\frac{z \sigma}{\sigma - 1} < 1$ where the last inequality follows from the assumption of decreasing returns to labor. We have characterized now the LHS as monotonically decreasing in $L$ and approaching $\frac{z \sigma}{\sigma - 1} < 1$ for large $L$.

Next, we show that the RHS is monotonically increasing in $L$. To see this we can express $C/Y = 1 - TC(W^*, \Delta)/Y$. From equations (??) and (??) we know $dC/dL > 0$, so it must be true that $Y$ increases faster than $TC$ when the size of the marker increases. This implies that when $L$ increases, $TC/Y$ decreases eventually reaching 0 when $L$ approaches infinity. This implies that the RHS increases as a function of $L$ and reaches 1 for very large $L$.

Given that LHS is monotonically decreasing in $L$ and RHS is monotonically increasing with $L$, then LHS and RHS intersect if and only if $LHS|_{L=0} > RHS|_{L=0}$. To calculate $LHS|_{L=0}$ simply set $s = 0$ to find $LHS|_{L=0} = \frac{1 - z}{\sigma - 1} \frac{6z \sigma}{16(z \sigma - 1) + 3(\sigma - 1)} + \frac{z \sigma}{\sigma - 1}$. To calculate $RHS|_{L=0}$ we first need to recall that $TC(W^*, \Delta) = \frac{2}{3} \frac{\pi \Delta^3}{(\mu gd)^2} s^3$ which from (19) is equal to $TC(W^*, \Delta) = \frac{2}{3} \frac{s W^S}{1 - 2s/3}$. Now, using $p^W/p^C = (1 - z)Y/W^S$ and from the definition $p^W/p^C = s/1 - s$ we find $C/Y = 1 - \frac{2}{3} \frac{(1 - s)(1 - z)}{1 - 2/3s}$. Hence when $L$ approaches zero, $RHS|_{L=0} = \frac{1 + 2z}{3} < 1$. Putting these two result together, there will be an intersection, and this intersection will be unique when $\sigma < 3/(1 - z)$. This is the Increasing Returns Case. If $\sigma > 3/(1 - z)$ there is no intersection and $C/L$ is monotonically decreasing with the size of the market. This is the Almost Neoclassical case.

The proof for production per capita is similar. Production is equal to $Y = \frac{\alpha}{\beta} (\sigma - 1) \left( \frac{1}{\alpha \sigma} L^z (W^S)^{1-z} \right)^{\sigma/(\sigma - 1)}$. The rate of change of production per capita can be written as

\[
\frac{\hat{Y}}{\hat{L}} = \frac{1 - (1 - z)\sigma}{\sigma - 1} \frac{\hat{L}}{\hat{L}} + \frac{(1 - z)\sigma}{\sigma - 1} W^* \tag{A.13}
\]
We know \( s \) approaches zero as \( L \) approaches zero, so production per capita first raises with the size of the market when the previous equation is positive, that is when \( \sigma < 3/(1-z) \). For large \( L \) production per capita always falls because as \( s \) approaches 1, the expression \( \frac{\tilde{Y}}{\bar{L}} \) approaches \( \frac{1-(1-z)\sigma}{\sigma-1} \) which is negative, assuming decreasing returns to labor in equilibrium. If \( \sigma > 3/(1-z) \), then production per capita falls monotonically with the size of the market.

**Proposition 3:** Equating (A.1) setting \( \hat{L} = 0 \) and rearranging terms we find

\[
\frac{\left( \frac{L^\omega}{w} \right)}{\Delta} = -\frac{3}{\frac{6(1-s)^2 z\sigma - 1}{\sigma-1} + 1} < 0 \quad (A.14)
\]

Replacing (A.14) in the demand equation in (A.1) we find \( dW_D/d\Delta > 0 \). Rearranging the expression in (A.10) and, given \( W^S = W^D \) in equilibrium, we find \( dC/d\Delta > 0 \).

Given the characteristics of \( LHS \) and \( RHS \) in equation (A.12), if \( LHS \) increases with \( \Delta \) and \( RHS \) decreases with \( \Delta \) for all \( L \), then it is necessarily true that \( L \) increases with \( \Delta \).

We start calculating the result of the \( LHS \):

\[
\frac{dLHS}{d\Delta} = -\frac{12z\sigma(1-z)(2-s)(1-s)}{(3\sigma + 2z\sigma - 3) + 2(7 - \sigma - 6z\sigma)s + 6(z\sigma - 1)s^2} \frac{ds}{d\Delta} > 0 \quad (A.15)
\]

where \( ds/d\Delta < 0 \) directly from equations (A.1)-(A.4). Similarly, for the \( RHS \) we have

\[
\frac{dRHS}{d\Delta} = \frac{2(1-z)}{(3-2s)^2} \frac{ds}{d\Delta} < 0 \quad (A.16)
\]

Then peak value of \( L \) increases with \( \Delta \). To show that the peak of consumption per capita increases with \( \Delta \) we simply need to show that \( C/L \) is an increasing function of \( \Delta \) for all \( L \); that is \( \frac{d(C/L)}{d\Delta} = \frac{dC}{d\Delta L} - \frac{C}{L^2} \frac{dL}{d\Delta} \) were the second term is equal to zero, so that \( \frac{d(C/L)}{d\Delta} = \frac{dC}{d\Delta L} > 0 \) directly from Proposition 4.

**Proposition 5:** The second order conditions required for a maximum imply \( \epsilon'_{g\rho} < \epsilon'_{h\rho} \) where we have defined \( \epsilon'_{h\rho} = -\frac{\rho h''(\rho)}{h'(\rho)} > 0 \) and \( \epsilon'_{g\rho} = -\frac{\rho g''(\rho)}{g'(\rho)} > 0 \).

To find how the optimal amount of infrastructure changes when power density increases take equation the first order condition for optimal investment and total differentiate it to find

\[
(\epsilon'_{h\rho} - \epsilon'_{g\rho}) \frac{\dot{\rho}}{\Delta} = -3 - \frac{9 - 8s}{3-2s} \frac{\dot{s}}{\Delta} \quad (A.17)
\]

Then, we can calculate the equilibrium price response to a change in \( \Delta \) while considering
changes in infrastructure.

\[
\frac{\hat{s}}{\Delta} = -\frac{3(1 - s)^{\frac{\sigma - 1}{\sigma - 1}}}{6(1-s)^2 \frac{3\sigma - 1}{\sigma - 1} + 1} + \frac{(1 - s)^{\frac{\sigma - 1}{\sigma - 1}} \epsilon_{g\rho} \hat{\rho}}{6(1-s)^2 \frac{3\sigma - 1}{\sigma - 1} + 1 \Delta} \tag{A.18}
\]

where \( \epsilon_{g\rho} = -\frac{\rho g'(\rho)}{g(\rho)} > 0 \). Then we find:

\[
\frac{\hat{\rho}}{\Delta} = -\frac{3}{\left(\epsilon'_{h\rho} - \epsilon'_{g\rho} + \frac{(9-8s)(1-s)^{\frac{\sigma - 1}{\sigma - 1}} \epsilon_{g\rho}}{6(1-s)^2 \frac{3\sigma - 1}{\sigma - 1} + 1}\right)} + \frac{\frac{3(9-8s)(1-s)^{\frac{\sigma - 1}{\sigma - 1}}}{6(1-s)^2 \frac{3\sigma - 1}{\sigma - 1} + 1}}{\left(\epsilon'_{h\rho} - \epsilon'_{g\rho} + \frac{(9-8s)(1-s)^{\frac{\sigma - 1}{\sigma - 1}} \epsilon_{g\rho}}{6(1-s)^2 \frac{3\sigma - 1}{\sigma - 1} + 1}\right)} \tag{A.19}
\]

where the denominator is positive directly from the second order conditions. The expression \( \frac{\hat{\rho}}{\Delta} < 0 \) if:

\[-1 < \left((9 - 8s)(1 - s) + 6(1 - s)^2\right)\frac{\epsilon_{g\rho}}{\epsilon_{g\rho} + \frac{(9-8s)(1-s)^{\frac{\sigma - 1}{\sigma - 1}}}{6(1-s)^2 \frac{3\sigma - 1}{\sigma - 1} + 1}} \tag{A.20}
\]

which always holds because we have assume \( \sigma > 1/z \) for an equilibrium to exist. To show that prices fall when \( \Delta \) increases and \( \rho \) is chosen endogenously we take equation (A.18) and notice that for \( \frac{\hat{s}}{\Delta} \) to be negative we require:

\[
\frac{\hat{s}}{\Delta} < 0 \Rightarrow -3 < -\epsilon_{g\rho} \frac{\hat{\rho}}{\Delta} \tag{A.21}
\]

but by definition \( \epsilon_{g\rho} > 0 \) and we just showed \( \frac{\hat{\rho}}{\Delta} < 0 \) so the above condition always holds. Similarly we can show that energy supplied to the city always increases with power density and endogenous infrastructure. To see this, calculate the total derivative of supplied energy and see how it changes with \( \Delta \):

\[
\frac{\overline{W^S}}{\Delta} = 3 + \frac{6(1-s)}{3-2s} \frac{\hat{s}}{\Delta} - \epsilon_{g\rho} \frac{\hat{\rho}}{\Delta} \tag{A.22}
\]

Replacing equation (A.18) and simplifying we find that \( \frac{\overline{W^S}}{\Delta} > 0 \) when \( 3 > \epsilon_{g\rho} \frac{\hat{\rho}}{\Delta} \) which always holds because by again definition \( \epsilon_{g\rho} > 0 \) and we just showed \( \frac{\hat{\rho}}{\Delta} < 0 \).
B Data

Population: We have collected the population data from several sources. Our key population data is taken from the recent estimates by Broadberry et al. (2011). They combine county-level population estimates from Russell (1948, 53-54), Campbell (2008, 926) and Wrigley (2009, 721) for the benchmark years of 1086, 1290, 1377, and 1600. Domesday Book of 1086 and the poll tax returns data for 1377 are also used by the authors as cross-sectional evidence of the population estimates. Our second set of county-level population data comes from census data for 1831 and 1841 (Gatley, 2000).

Coal Fields: We measure the areas of coal fields for counties that have coal located within its borders. The Coal and Iron map by the Edinburgh Geographical Institute is used to locate the exposed and covered coal fields and measure their areas. The areas of coal fields are measured in square miles.

The data on county area, which is from 1831 census and also measured in square miles, is used together with the coal field areas data to construct the shares of exposed and covered coal fields in the area of counties.

Wool and Cloth Exports: The data on the amount of wool exports from ports in England is from Carus-Wilson and Coleman (1963, 36-73). The dataset covers the period between 1279 and 1546. The wool exports are measured in sacks of wool, the sack being a standard weight of 364 lb. of shorn wool.

The cloth exports by ports in England are also from Carus-Wilson and Coleman (1963, 75-119) and the data available covers the period between 1348 and 1546. The unit of measurement for cloth exports is ‘cloth of assize’ which is approximately 24 yards long and 1.5-2 yards wide. For customs purposes, cloths with different measurements were converted into ‘cloth of assize’ in port records.